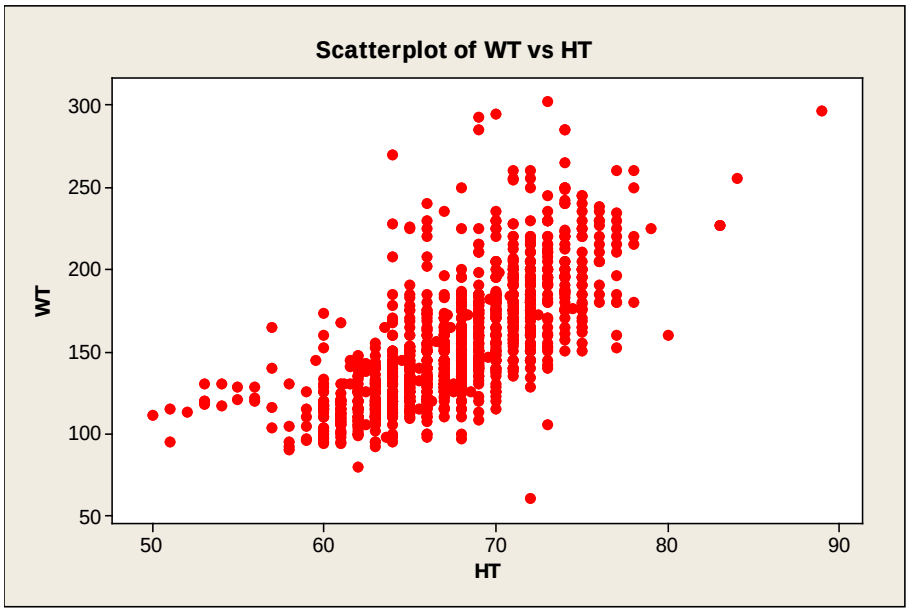
Example – What is the relationship between height and weight for UF students?

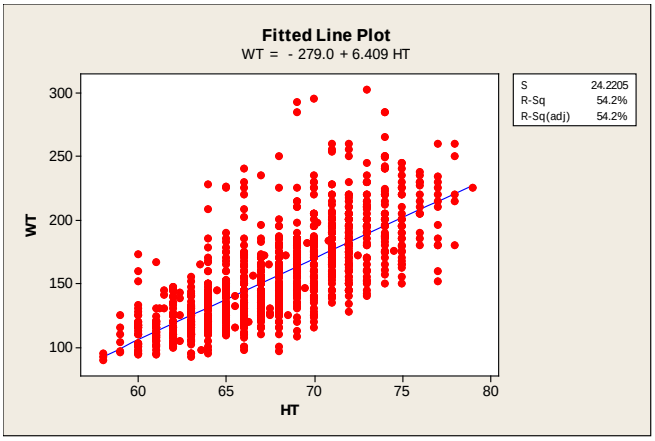
Data on UF students' heights and weights collected by STA3024 students. N=1309

Questions about some data – are these heights correct?

	HT	WT
F	50.0	111
F	51.0	115
F	51.0	95
F	52.0	113
F	53.0	118
F	53.0	120
F	53.0	120
F	53.0	130
F	54.0	117
F	54.0	130
F	55.0	121
F	55.0	128
F	56.0	120
F	56.0	122
F	56.0	128
F	57.0	103
F	57.0	116
F	57.0	140
M	57.0	165
F	58.0	104
F	58.0	130
F	58.0	90
F	58.0	92
F	58.0	95
F	59.0	104
F	59.0	110
F	59.0	115
F	59.0	125
F	59.0	96
F	59.0	97
F	59.5	145
M	80	160
M	83	227
M	83	227
M	84	255
M	89	296
M	72	60
M	73	105
F	64	270



Regression Analysis: WT versus HT



The regression equation is
WT = - 279 + 6.41 HT

Predictor	Coef	SE Coef	T	P
Constant	-279.01	11.19	-24.92	0.000
HT	6.4088	0.1649	38.86	0.000

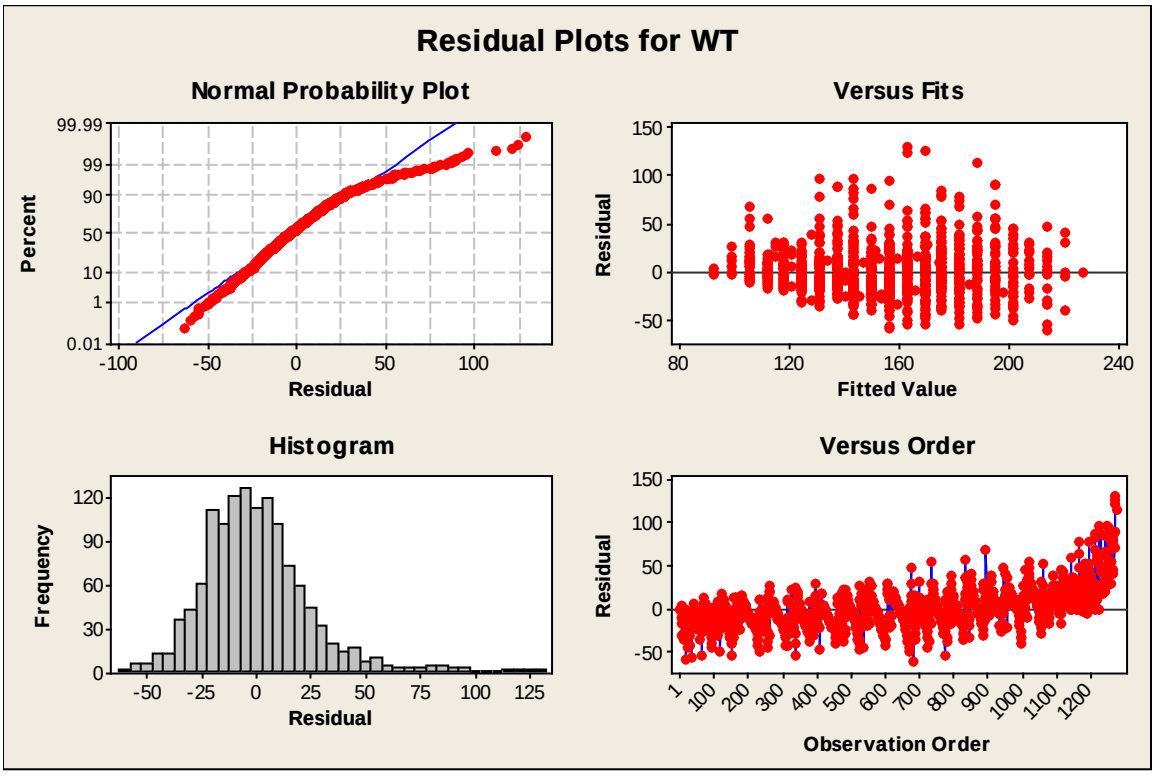
S = 24.2205 R-Sq = 54.2% R-Sq(adj) = 54.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	885986	885986	1510.29	0.000
Residual Error	1276	748543	587		
Total	1277	1634529			

Predicted Values for New Observations

New Obs	HT	Fit	SE Fit	95% CI	95% PI
1	65	137.562	0.816	(135.961, 139.163)	(90.019, 185.106)
2	60	105.518	1.448	(102.678, 108.359)	(57.917, 153.120)
3	76	208.059	1.519	(205.080, 211.038)	(160.449, 255.669)



Regression Analysis: WT_F versus HT_F

```
The regression equation is
WT_F = - 125 + 3.96 HT_F

Predictor     Coef  SE Coef      T      P
Constant   -125.21    17.53  -7.14  0.000
HT_F        3.9614   0.2700  14.67  0.000

S = 19.1292   R-Sq = 24.9%   R-Sq(adj) = 24.8%

Analysis of Variance

Source           DF      SS     MS       F      P
Regression        1   78781  78781  215.29  0.000
Residual Error  650  237852    366
Total           651  316633
```

Regression Analysis: WT_M versus HT_M

```
The regression equation is
WT_M = - 184 + 5.14 HT_M

Predictor     Coef  SE Coef      T      P
Constant   -184.21    25.73  -7.16  0.000
HT_M        5.1421   0.3633  14.16  0.000

S = 26.5446   R-Sq = 24.3%   R-Sq(adj) = 24.2%

Analysis of Variance

Source           DF      SS      MS       F      P
Regression        1  141187  141187  200.37  0.000
Residual Error  624  439681     705
Total           625  580868
```

Regression Analysis: WT versus HT, GENDER_M_1

```
The regression equation is
WT = - 165 + 4.57 HT + 21.0 GENDER_M_1

Predictor      Coef  SE Coef       T      P
Constant    -164.68    14.76  -11.16  0.000
HT           4.5699   0.2271   20.12  0.000
GENDER_M_1   20.963    1.866   11.23  0.000

S = 23.1134   R-Sq = 58.3%   R-Sq(adj) = 58.3%

Analysis of Variance

Source            DF       SS      MS       F      P
Regression         2   953389  476695  892.31  0.000
Residual Error  1275   681140     534
Total           1277  1634529
```

Example: Predicting College GPA – data from book

Regression Analysis: CGPA versus Height, Gender, etc

```
The regression equation is
CGPA = 0.53 + 0.0194 Height + 0.047 Gender - 0.00163 Haircut - 0.042 Job
       + 0.0004 Studytime - 0.375 Smokecig + 0.0488 Dated + 0.546 HSGPA
       + 0.00315 HomeDist + 0.00069 BrowseInternet - 0.00128 WatchTV
       - 0.0117 Exercise + 0.0140 ReadNewsP + 0.039 Vegan
       - 0.0139 PoliticalDegree - 0.0801 PoliticalAff


Predictor              Coef    SE Coef      T      P
Constant              0.532      1.496   0.36  0.724
Height              0.01942    0.01637   1.19  0.242
Gender               0.0468     0.1429   0.33  0.745
Haircut           -0.001633   0.001697  -0.96  0.341
Job                 -0.0418     0.1024  -0.41  0.685
Studytime           0.00043    0.01921   0.02  0.982
Smokecig            -0.3746     0.2249  -1.67  0.103
Dated               0.04881    0.07111   0.69  0.496
HSGPA                0.5457     0.1776   3.07  0.004
HomeDist           0.003147   0.003400   0.93  0.360
BrowseInternet     0.000689   0.001163   0.59  0.557
WatchTV          -0.0012840  0.0009710  -1.32  0.193
Exercise          -0.011657   0.005934  -1.96  0.056
ReadNewsP           0.01395    0.02272   0.61  0.543
Vegan                0.0392     0.1578   0.25  0.805
PoliticalDegree    -0.01390    0.03185  -0.44  0.665
PoliticalAff       -0.08006    0.07741  -1.03  0.307


S = 0.322198   R-Sq = 43.2%   R-Sq(adj) = 21.5%


Analysis of Variance

Source          DF      SS      MS     F      P
Regression      16  3.3135  0.2071  1.99  0.037
Residual Error  42  4.3601  0.1038
Total           58  7.6736


Unusual Observations

Obs  Height    CGPA     Fit  SE Fit  Residual  St Resid
 28    67.0  2.9800  3.5898  0.2442   -0.6098     -2.90R
 40    65.0  3.9300  3.3458  0.2176    0.5842      2.46R
 59    62.0  2.5000  3.4718  0.1352   -0.9718     -3.32R

R denotes an observation with a large standardized residual.
```

Best Subsets Regression: CGPA versus Height, Gender, ...

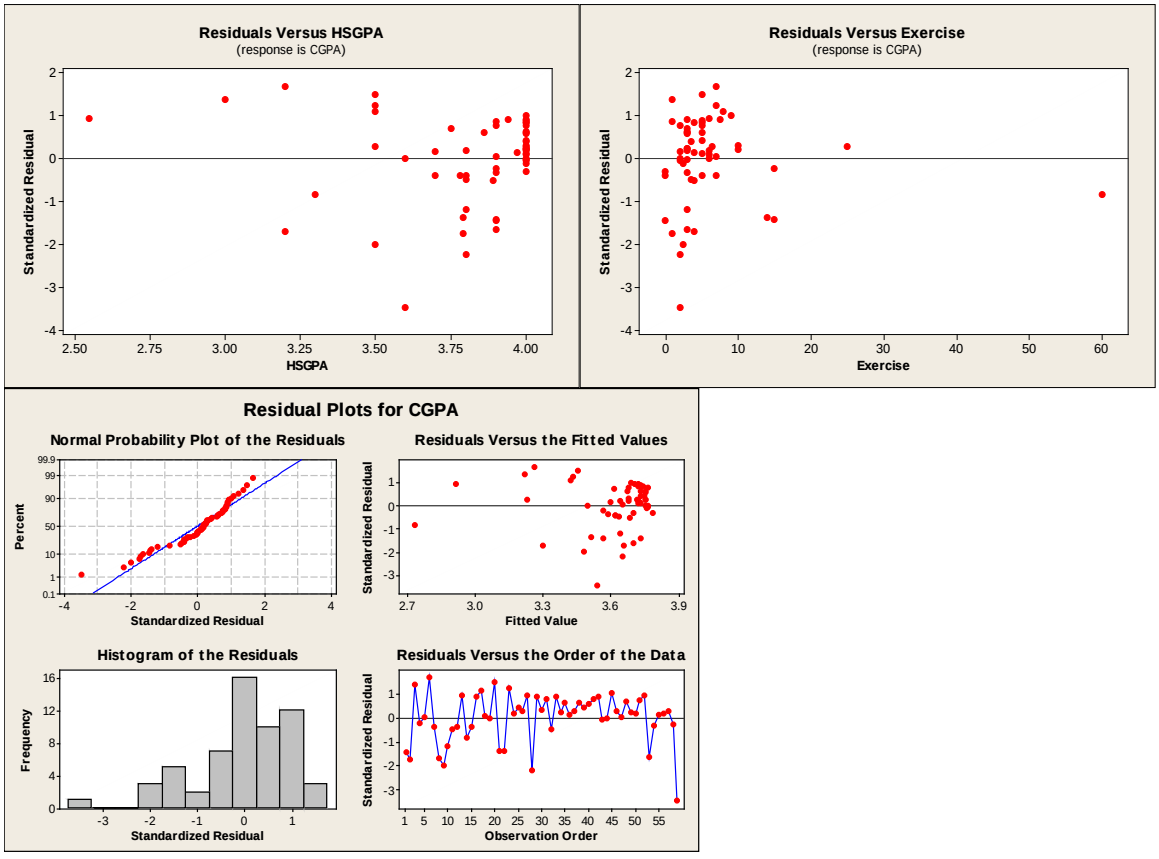
Response is CGPA

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	Height	Gender	Haircut	Job	StudyTime	Smoking	Date	HSGPA	HomeDist	BrowseInternet	WatchTV	Exercise	ReadNewsP	Vegan	PoliticalDegree	PoliticalAff
1	25.5	24.2	0.1	0.31667								X								
1	13.0	11.5	9.3	0.34217												X				
2	31.6	29.2	-2.4	0.30613								X				X				
2	29.4	26.9	-0.8	0.31109								X			X					
3	33.8	30.2	-2.1	0.30389			X					X				X				
3	33.7	30.0	-2.0	0.30423								X			X	X				
4	35.7	31.0	-1.5	0.30223						X		X			X	X				
4	35.3	30.5	-1.2	0.30320			X					X			X	X				
5	37.3	31.4	-0.6	0.30132	X					X		X			X	X				
5	37.0	31.1	-0.4	0.30198			X			X		X			X	X				
6	38.3	31.2	0.6	0.30163	X		X			X		X			X	X				
6	38.3	31.2	0.6	0.30164	X					X		X			X	X				X
7	39.6	31.3	1.7	0.30150	X		X			X		X			X	X				X
7	39.3	30.9	1.9	0.30231	X					X		X	X		X	X				X
8	40.4	30.8	3.1	0.30249	X		X			X		X			X	X			X	X
8	40.4	30.8	3.1	0.30256	X		X			X		X	X		X	X				X
9	41.5	30.8	4.2	0.30266	X		X			X		X	X		X	X			X	X
9	41.0	30.2	4.6	0.30395	X		X	X		X		X	X		X	X				X
10	41.9	29.8	6.0	0.30478	X		X			X	X	X	X		X	X			X	X
10	41.8	29.7	6.0	0.30492	X		X	X		X		X	X		X	X			X	X
11	42.2	28.7	7.7	0.30712	X		X	X		X	X	X	X		X	X			X	X
11	42.2	28.7	7.7	0.30715	X		X			X	X	X	X		X	X	X		X	X
12	42.6	27.6	9.4	0.30945	X		X			X	X	X	X	X	X	X	X		X	X
12	42.6	27.6	9.5	0.30954	X		X	X		X	X	X	X	X	X	X			X	X
13	42.9	26.4	11.2	0.31205	X		X	X		X	X	X	X	X	X	X	X		X	X
13	42.8	26.3	11.3	0.31229	X	X	X			X	X	X	X	X	X	X	X		X	X
14	43.1	25.0	13.1	0.31502	X	X	X	X		X	X	X	X	X	X	X	X		X	X
14	43.0	24.9	13.1	0.31526	X		X	X		X	X	X	X	X	X	X	X	X	X	X
15	43.2	23.4	15.0	0.31843	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X
15	43.1	23.2	15.1	0.31866	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X
16	43.2	21.5	17.0	0.32220	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Regression Analysis: CGPA versus HSGPA, Exercise

```
The regression equation is
CGPA = 1.55 + 0.560 HSGPA - 0.0111 Exercise

Predictor       Coef   SE Coef      T      P
Constant      1.5489    0.5551   2.79  0.007
HSGPA         0.5599    0.1436   3.90  0.000
Exercise   -0.011138  0.004985  -2.23  0.029

S = 0.306126   R-Sq = 31.6%   R-Sq(adj) = 29.2%

Analysis of Variance
Source          DF      SS      MS      F      P
Regression       2  2.4256  1.2128  12.94  0.000
Residual Error  56  5.2479  0.0937
Total           58  7.6736

Unusual Observations
Obs  HSGPA    CGPA     Fit  SE Fit  Residual  St Resid
  3   3.00  3.6000  3.2176  0.1297    0.3824      1.38 X
  9   3.50  2.8800  3.4808  0.0642   -0.6008     -2.01R
 14   3.30  2.6000  2.7284  0.2647   -0.1284     -0.83 X
 27   2.55  3.1400  2.9099  0.1840    0.2301      0.94 X
 28   3.80  2.9800  3.6544  0.0445   -0.6744     -2.23R
 59   3.60  2.5000  3.5424  0.0556   -1.0424     -3.46R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.
```

Residual Plots for CGPA



Regression Analysis: CGPA versus HSGPA, Exercise

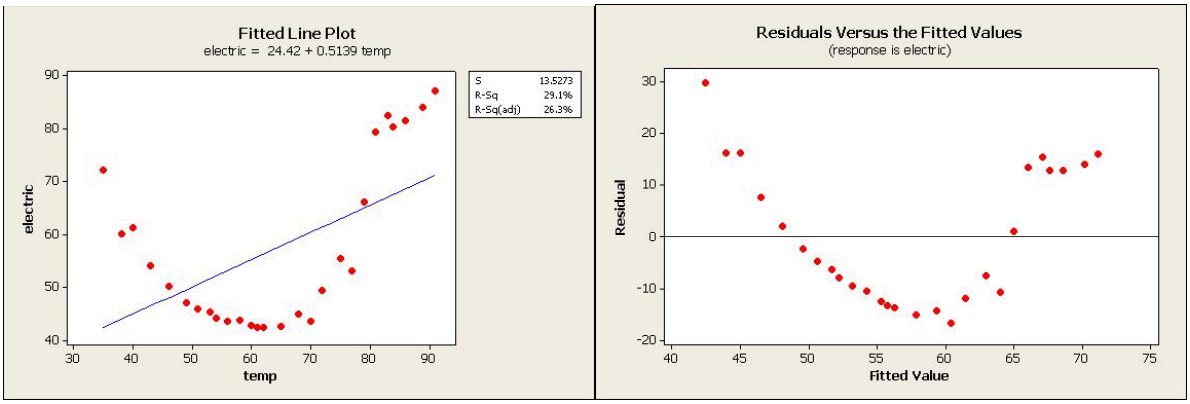
```
The regression equation is
CGPA = 1.54 + 0.554 HSGPA - 0.00432 Exercise

Predictor       Coef   SE Coef      T      P
Constant      1.5388    0.5568   2.76  0.008
HSGPA         0.5542    0.1441   3.85  0.000
Exercise   -0.004320  0.009596  -0.45  0.654

S = 0.306969   R-Sq = 21.9%   R-Sq(adj) = 19.0%

Analysis of Variance

Source          DF       SS       MS     F      P
Regression       2  1.45009  0.72504  7.69  0.001
Residual Error  55  5.18265  0.09423
Total           57  6.63274

Unusual Observations
Obs  HSGPA    CGPA     Fit  SE Fit  Residual  St Resid
  3   3.00  3.6000  3.1970  0.1324    0.4030      1.45 X
 25   3.50  3.3100  3.3705  0.1974   -0.0605     -0.26 X
 26   2.55  3.1400  2.9261  0.1856    0.2139      0.87 X
 27   3.80  2.9800  3.6361  0.0497   -0.6561     -2.17R
 58   3.60  2.5000  3.5252  0.0594   -1.0252     -3.40R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.
```

Regression Analysis: CGPA versus HSGPA

```
The regression equation is
CGPA = 1.50 + 0.560 HSGPA

Predictor    Coef  SE Coef     T      P
Constant   1.4964   0.5448  2.75  0.008
HSGPA      0.5596   0.1426  3.92  0.000

S = 0.304776   R-Sq = 21.6%   R-Sq(adj) = 20.2%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  1.4310  1.4310  15.41  0.000
Residual Error  56  5.2017  0.0929
Total           57  6.6327

Unusual Observations

Obs  HSGPA    CGPA     Fit  SE Fit  Residual  St Resid
  3   3.00  3.6000  3.1753  0.1223    0.4247      1.52 X
 26   2.55  3.1400  2.9234  0.1842    0.2166      0.89 X
 27   3.80  2.9800  3.6230  0.0400   -0.6430     -2.13R
 58   3.60  2.5000  3.5111  0.0500   -1.0111     -3.36R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.
```

Example: Electrical Consumption vs. Temperature

Linear Regression



Regression Analysis: electric versus temp

The regression equation is

electric = 24.4 + 0.514 temp

Predictor	Coef	SE Coef	T	P
Constant	24.42	10.57	2.31	0.029
temp	0.5139	0.1603	3.21	0.004

S = 13.5273 R-Sq = 29.1% R-Sq(adj) = 26.3%

Analysis of Variance

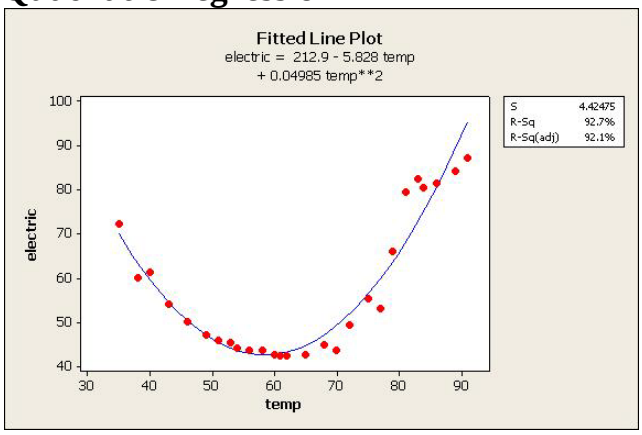
Source	DF	SS	MS	F	P
Regression	1	1880.7	1880.7	10.28	0.004
Residual Error	25	4574.7	183.0		
Total	26	6455.5			

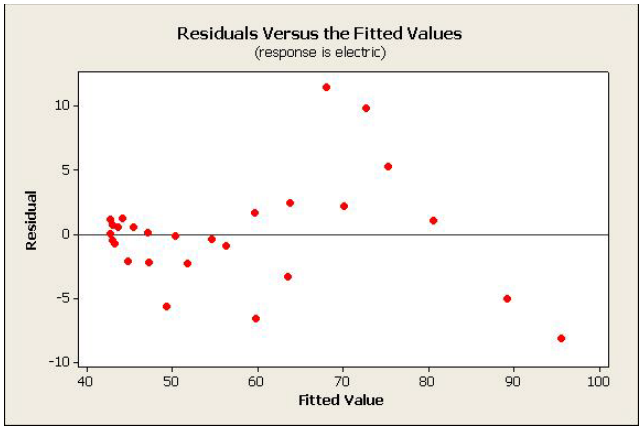
Unusual Observations

Obs	temp	electric	Fit	SE Fit	Residual	St Resid
1	35.0	72.16	42.40	5.32	29.76	2.39R

R denotes an observation with a large standardized residual.

Quadratic Regression





Regression Analysis: electric versus temp, temp2

The regression equation is
electric = 213 - 5.83 temp + 0.0499 temp**2

Predictor	Coef	SE Coef	T	P
Constant	212.93	13.47	15.81	0.000
temp	-5.8278	0.4411	-13.21	0.000
temp**2	0.049854	0.003443	14.48	0.000

S = 4.42475 R-Sq = 92.7% R-Sq(adj) = 92.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	5985.6	2992.8	152.86	0.000
Residual Error	24	469.9	19.6		
Total	26	6455.5			

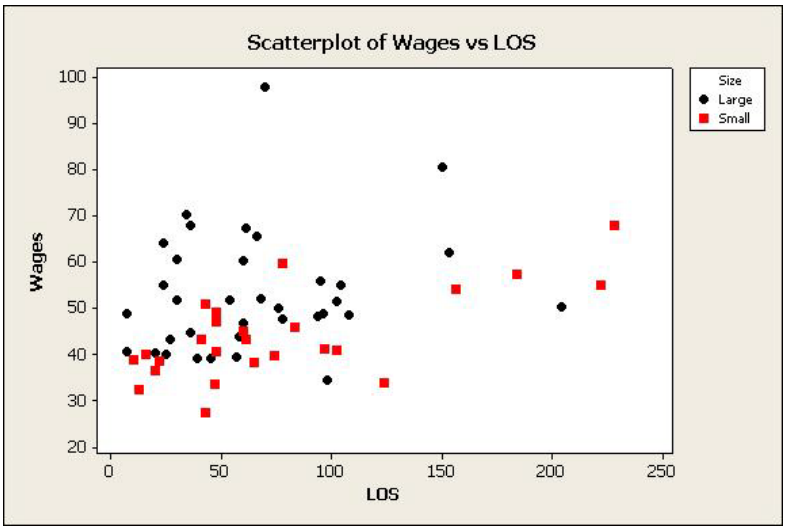
Source	DF	Seq SS
temp	1	1880.7
temp2	1	4104.8

Unusual Observations

Obs	temp	electric	Fit	SE Fit	Residual	St Resid
1	35.0	72.164	70.032	2.582	2.132	0.59 X
22	81.0	79.468	67.974	1.243	11.494	2.71R
23	83.0	82.469	72.671	1.369	9.798	2.33R
27	91.0	87.265	95.445	2.356	-8.180	-2.18R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.

Example: Wages vs Length of Service and Size of Company



Coding of size of company: small = 0 large = 1

Regression Analysis: Wages versus LOS, size, LOS*size

The regression equation is
Wages = 35.9 + 0.104 LOS + 13.6 size - 0.0483 LOS*size

Predictor	Coef	SE Coef	T	P
Constant	35.914	3.562	10.08	0.000
LOS	0.10424	0.03632	2.87	0.006
size	13.631	4.910	2.78	0.007
LOS*size	-0.04828	0.05634	-0.86	0.395

S = 10.9612 R-Sq = 26.6% R-Sq(adj) = 22.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	2438.1	812.7	6.76	0.001
Residual Error	56	6728.3	120.1		
Total	59	9166.4			

Source	DF	Seq SS
LOS	1	843.5
size	1	1506.3
LOS*size	1	88.2

Regression Analysis: Wages versus LOS, size

The regression equation is
Wages = 37.5 + 0.0842 LOS + 10.2 size

Predictor	Coef	SE Coef	T	P
Constant	37.466	3.061	12.24	0.000
LOS	0.08417	0.02770	3.04	0.004
size	10.228	2.882	3.55	0.001

S = 10.9357 R-Sq = 25.6% R-Sq(adj) = 23.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	2349.9	1174.9	9.82	0.000
Residual Error	57	6816.6	119.6		
Total	59	9166.4			

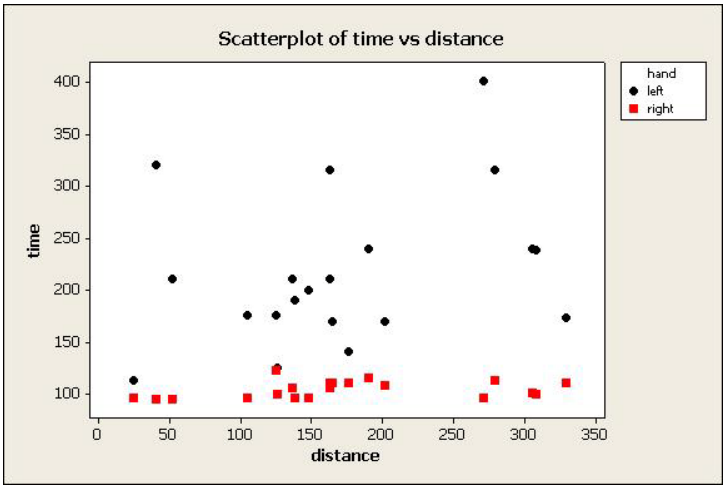
Source	DF	Seq SS
LOS	1	843.5
size	1	1506.3

Unusual Observations

Obs	LOS	Wages	Fit	SE Fit	Residual	St Resid
15	70	97.68	53.59	1.85	44.09	4.09R
22	222	54.95	56.15	4.57	-1.21	-0.12 X
29	98	34.34	55.94	2.05	-21.60	-2.01R
42	228	67.91	56.66	4.71	11.25	1.14 X
47	204	50.17	64.87	4.26	-14.69	-1.46 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.

Example: Reaction Time in a Computer Game vs Distance to move mouse and Hand used.



Coding of hand: right = 0 left = 1

Regression Analysis: time versus distance, hand, dist*hand

The regression equation is
time = 99.4 + 0.028 distance + 72.2 hand + 0.234 dist*hand

Predictor	Coef	SE Coef	T	P
Constant	99.36	25.25	3.93	0.000
distance	0.0283	0.1308	0.22	0.830
hand	72.18	35.71	2.02	0.051
dist*hand	0.2336	0.1850	1.26	0.215

S = 50.6067 R-Sq = 59.8% R-Sq(adj) = 56.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	136948	45649	17.82	0.000
Residual Error	36	92198	2561		
Total	39	229146			

Source	DF	Seq SS
distance	1	6303
hand	1	126562
dist*hand	1	4083

Unusual Observations

Obs	distance	time	Fit	SE Fit	Residual	St Resid
25	163	315.00	214.29	11.38	100.71	2.04R
30	271	401.00	242.65	17.19	158.35	3.33R
31	40	320.00	182.09	20.68	137.91	2.99R

R denotes an observation with a large standardized residual.

Regression Analysis: time versus distance, hand

The regression equation is
time = 79.2 + 0.145 distance + 112 hand

Predictor	Coef	SE Coef	T	P
Constant	79.21	19.72	4.02	0.000
distance	0.14512	0.09324	1.56	0.128
hand	112.50	16.13	6.97	0.000

S = 51.0116 R-Sq = 58.0% R-Sq(adj) = 55.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	132865	66433	25.53	0.000
Residual Error	37	96281	2602		
Total	39	229146			

Unusual Observations

Obs	distance	time	Fit	SE Fit	Residual	St Resid
25	163	315.00	215.39	11.44	99.61	2.00R
30	271	401.00	231.10	14.67	169.90	3.48R
31	40	320.00	197.55	16.80	122.45	2.54R

R denotes an observation with a large standardized residual.

Regression Analysis: time versus hand

The regression equation is
time = 104 + 112 hand

Predictor	Coef	SE Coef	T	P
Constant	104.25	11.62	8.97	0.000
hand	112.50	16.43	6.85	0.000

S = 51.9573 R-Sq = 55.2% R-Sq(adj) = 54.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	126562	126562	46.88	0.000
Residual Error	38	102583	2700		
Total	39	229146			

Unusual Observations

Obs	hand	time	Fit	SE Fit	Residual	St Resid
30	1.00	401.00	216.75	11.62	184.25	3.64R
31	1.00	320.00	216.75	11.62	103.25	2.04R
32	1.00	113.00	216.75	11.62	-103.75	-2.05R

R denotes an observation with a large standardized residual.

One-way ANOVA: time versus hand

```
Source  DF      SS      MS      F      P
hand     1  126563  126563  46.88  0.000
Error   38  102584    2700
Total   39  229146

S = 51.96   R-Sq = 55.23%   R-Sq(adj) = 54.05%

                          Individual 95% CIs For Mean Based on
                          Pooled StDev
Level   N    Mean  StDev   +---------+---------+---------+---------
0      20  104.25   8.25   (-----*-----)
1      20  216.75  73.01                              (-----*-----)
                          +---------+---------+---------+---------
                          80        120       160       200
Pooled StDev = 51.96
```

Two-Sample T-Test and CI: time, hand

```
Two-sample T for time

hand   N    Mean  StDev  SE Mean
0     20  104.25   8.25      1.8
1     20   216.8   73.0       16

Difference = mu (0) - mu (1)
Estimate for difference:  -112.500
95% CI for difference:  (-146.889, -78.111)
T-Test of difference = 0 (vs not =): T-Value = -6.85  P-Value = 0.000
DF = 19
```